

# Topological perturbation of complex networks

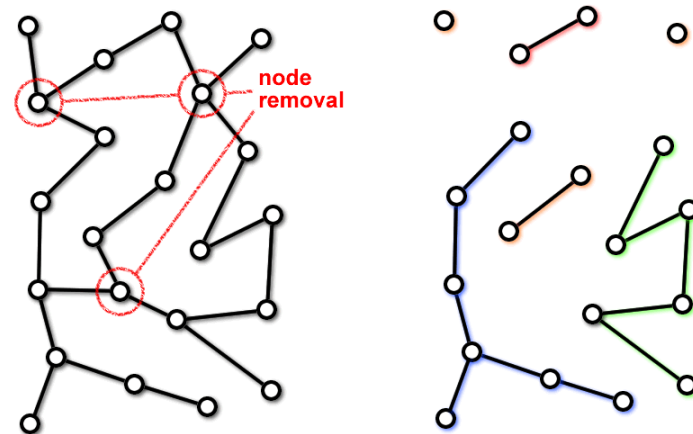
Perturbations in complex systems can deactivate some of the edges or nodes.

Edge loss: the edge is deleted

Node loss: the node and all its edges are deleted

Effects on the global topology:

- increase of path lengths,
- separation into isolated clusters.



More connected network - less effect of an edge removal

But bridges are definite points of vulnerability!

The effect of a node removal depends on the number and characteristics of its edges.

# Resilience to perturbations

Topological resilience studied in the literature:

the remaining nodes are still connected.

the average distance does not increase.

Ex. Propose other measures of resilience.

Testing resilience to incremental damage:

remove edges/nodes one by one, and look at

- the size of the giant connected component
- the average distance between nodes in the giant connected component

Ex. What factors affect the topological resilience of a network?

# Edge removal in random graphs

Start with a connected ER random graph with conn. prob.  $p$ .

$$p > \ln N / N$$

Remove a random fraction  $f$  of the edges.

Expected result: an ER graph with conn. prob.  $p(1-f)$

Connected if  $p(1-f) > \ln N / N$

For a broad class of starting graphs, there exists a threshold edge removal probability such that if a smaller fraction of edges is removed the graph is still connected.

B. Bollobas, Random Graphs, 1985

# Node removal in general random graphs

Consider a random graph with arbitrary  $P(k_0)$

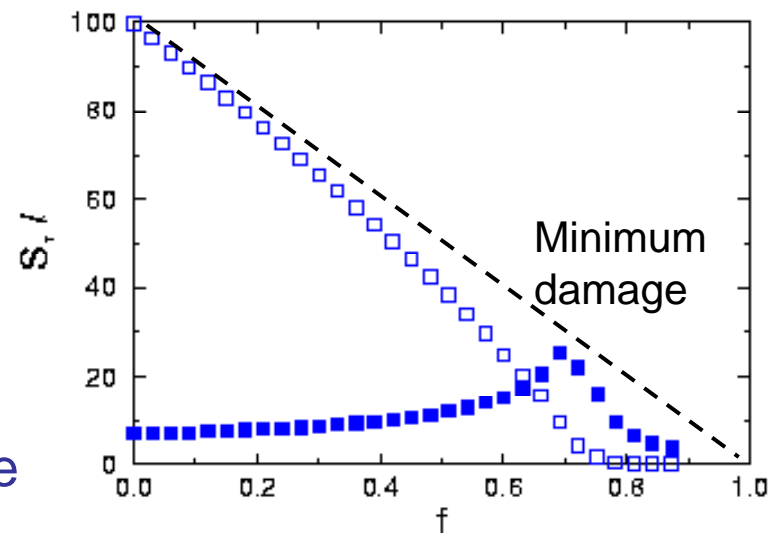
By a similar argument as in edge removal, the graph still has a giant connected component after a fraction  $f$  of nodes were removed if

$$f < f_c = 1 - \frac{1}{\frac{\langle k_0^2 \rangle}{\langle k_0 \rangle} - 1}$$

In **ER random graphs**

$$f_c = 1 - \frac{1}{\langle k_0 \rangle}$$

The higher the original average degree, the larger damage the network can survive



Q: How do you explain the peak in the average distance?

Cohen et al., Phys. Rev. Lett. 85, 4626 (2000).

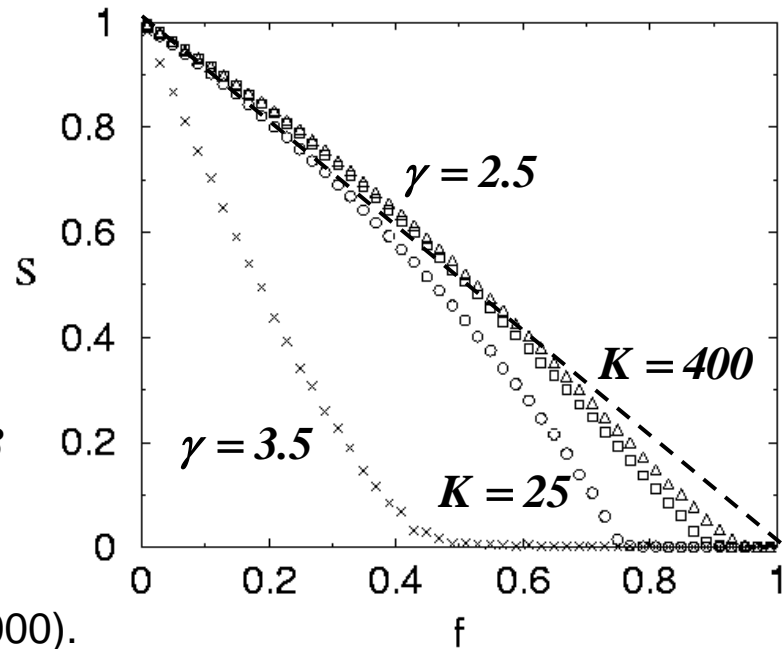
# Breakdown threshold of scale-free random graphs

Scale-free random graph with

$$P(k) = Ak^{-\gamma}, \text{ with } k = m, \dots, K$$

$$f_c = 1 - \frac{1}{\frac{\gamma-2}{\gamma-3}m-1} \quad \text{if } \gamma > 3$$

$$f_c = 1 - \frac{1}{\frac{\gamma-2}{3-\gamma}m^{\gamma-2}K^{3-\gamma}-1} \quad \text{if } 2 < \gamma < 3$$

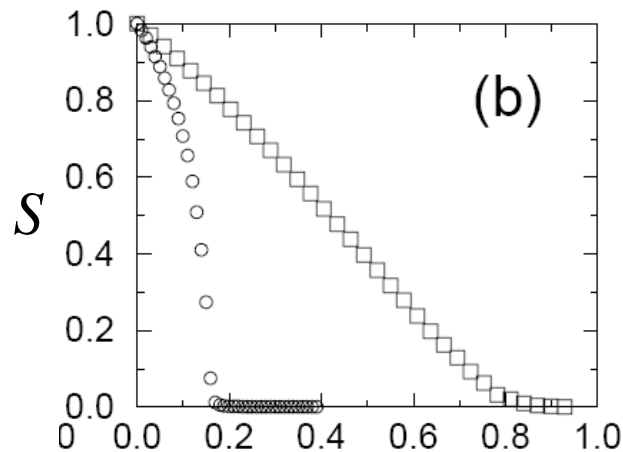


Cohen et al., Phys. Rev. Lett. 85, 4626 (2000).

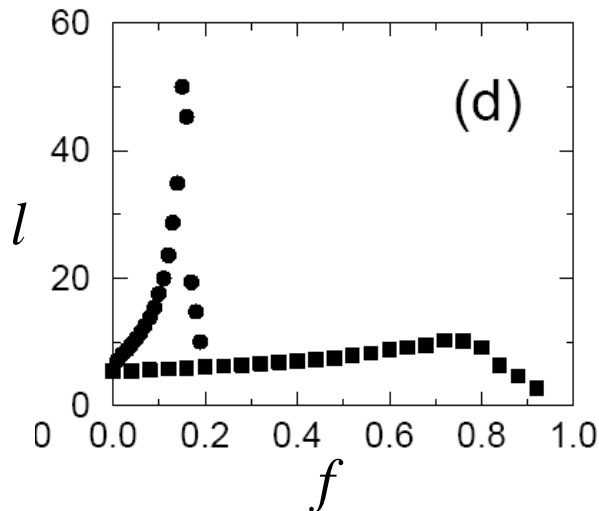
Infinite scale-free networks with  $\gamma < 3$  do not break down under random node failure.

Q. Do you think there is a flip side of this resilience to random node removal?

# Scale-free networks are error tolerant, but vulnerable to attacks



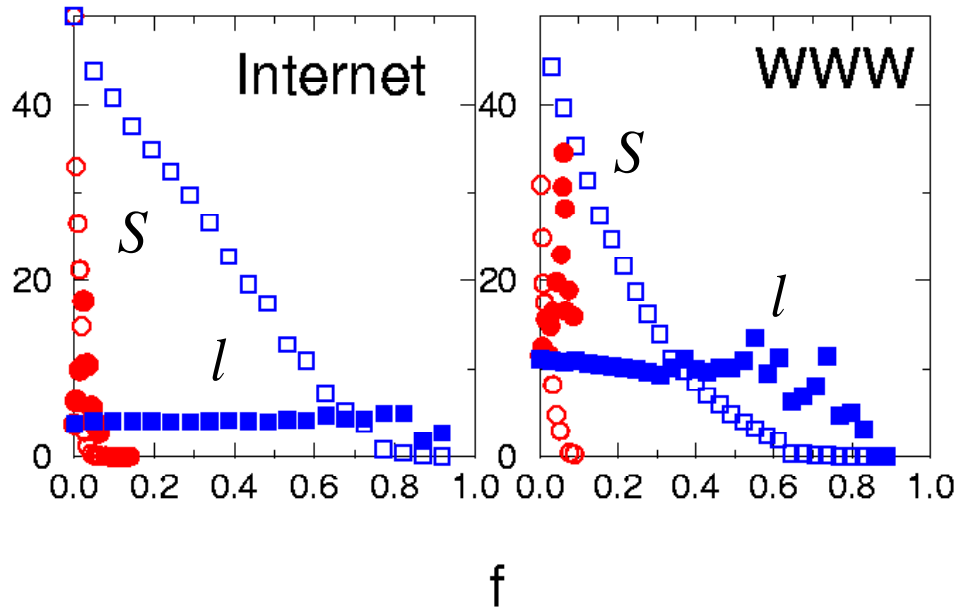
Network generated with the BA model  
(degree exponent of 3)  
Error: removal of a randomly selected node  
(squares)  
Attack: remove the node with the highest  
degree (circles)



**Errors:** little effect on the integrity of  
the network.  
**Attacks:** fast breakdown

R. Albert, H. Jeong, A.-L. Barabási,  
Nature 406, 378 (2000)

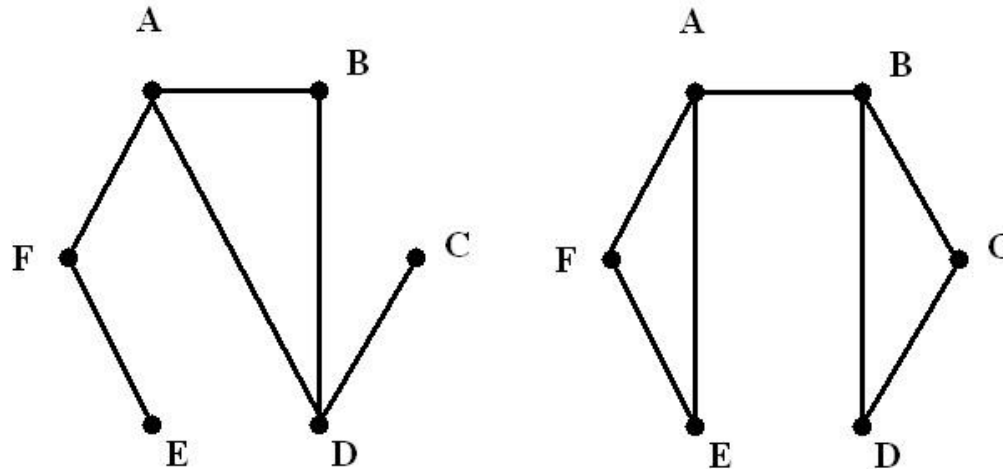
# Real scale-free networks show the same dual behavior



- blue squares: random failure
- red circles: targeted attack
- open symbols:  $S$
- filled symbols:  $l$

- break down if 5% of the nodes are eliminated selectively (always the highest degree node)
- resilient to the random failure of 50% of the nodes.

Similar results have been obtained for metabolic networks and food webs.



1. Rank order the nodes by your expectation for the effect of their removal. What were your criteria in doing so?
2. For each node, determine what is the effect of its removal on the size of the connected component.
3. Do the results match your expectations?



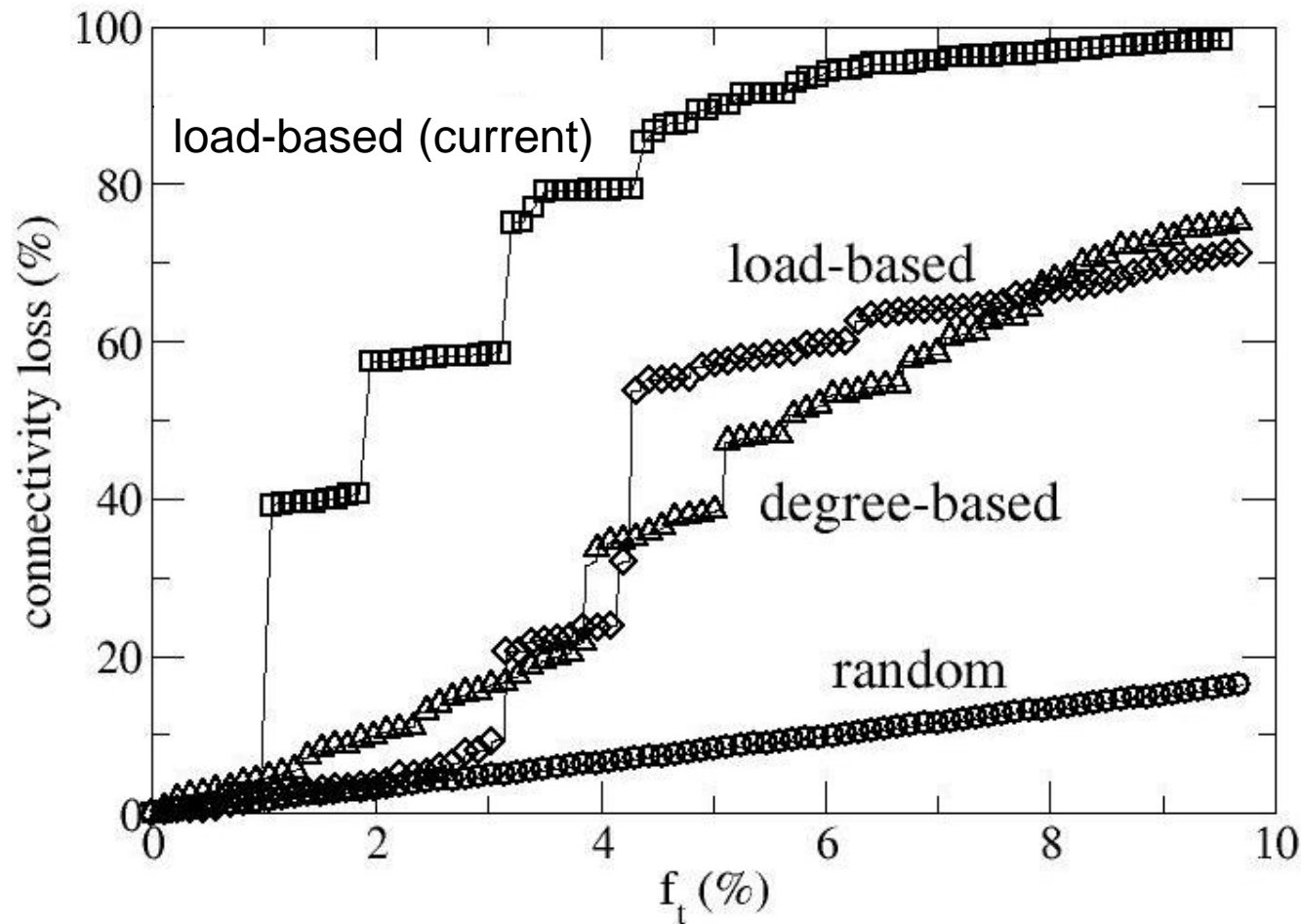
# Resilience of NA powergrid

- 14,000 nodes include generators, transmission substations, distribution substations
- 19,700 edges: high-voltage transmission lines
- The role of the power grid is to route power from generators to distribution substations (and then to customers)
- Connected network: power from any generator is in principle accessible to any substation, but 15% of edges are bridges.
- Connectivity measure used: the fraction of generators able to feed a given substation, averaged over substations

- Connectivity loss: 
$$CL = 1 - \left\langle \frac{N_g^i}{N_g} \right\rangle_i$$

R. Albert, I. Albert, G.N. Nakarado, Phys. Rev. E (2004)

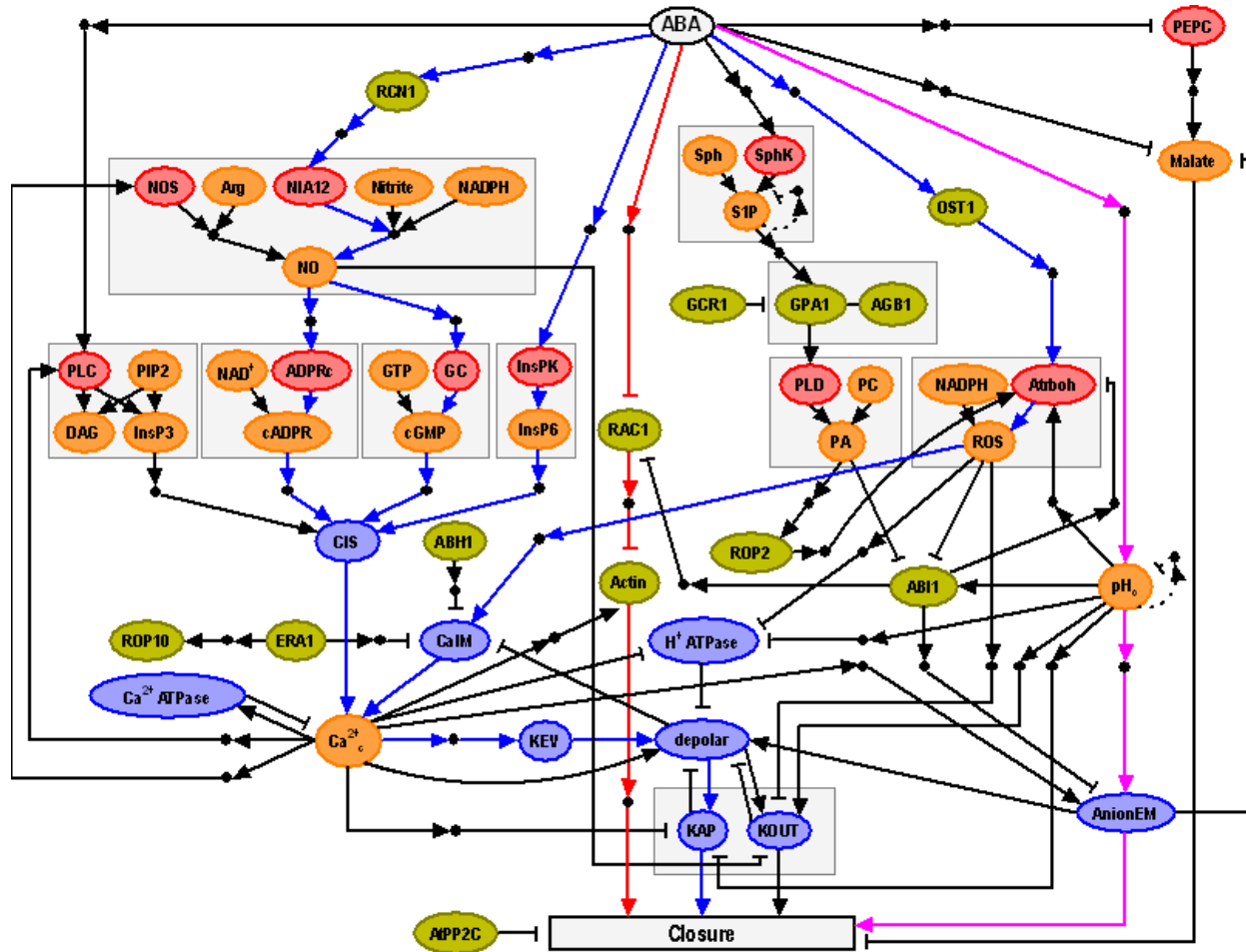
# Connectivity loss for transmission substation removal



Load =  
Betweenness  
centrality

Highest damage if the next substation to be removed is the current highest-betweenness substation.

# Resilience of the ABA signal transduction network



Q: What connectivity measures could be considered?

One possibility: the existence of at least one path from the source node (ABA) to the sink node (closure).

At least four separate ABA-closure paths, through  $\text{Ca}^{2+}$ , through actin, through  $\text{pH}_c$  and through malate.

4 nodes (e.g depolar, actin,  $\text{pH}_c$ , malate) need to be simultaneously disrupted to block all ABA- closure paths.

# Limitations of topological resilience

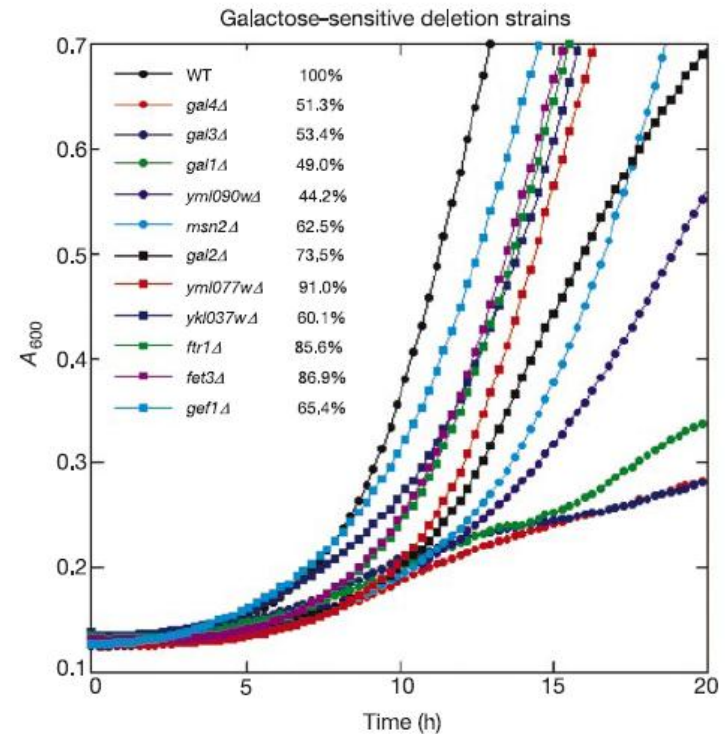
- The most relevant measure of connectivity may not be the size of the giant connected cluster
- The effects of removing a node or edge propagate through the network
  - E.g. cascading failure on the power grid, gene mutation
  - Depends on the dynamical properties of the network
- The network topology still determines the boundaries of propagating failure

# Resilience of cellular interaction networks

- Perturbation: knockout mutation of a gene. This means that all products of this gene (mRNA, protein) will be absent.
- Measured outcome: phenotype (e.g. growth behavior) of the mutant strain.
- The literature aims to correlate topological measures of the gene product (usually a protein in a protein interaction network) with the phenotype of the gene mutation.
- Caveats
  - The gene knockout may be incompletely represented by the loss of a protein node in a protein-protein interaction network
  - The effects of knockouts propagate through the network

# Systematic deletion of *S.cerevisiae* genes

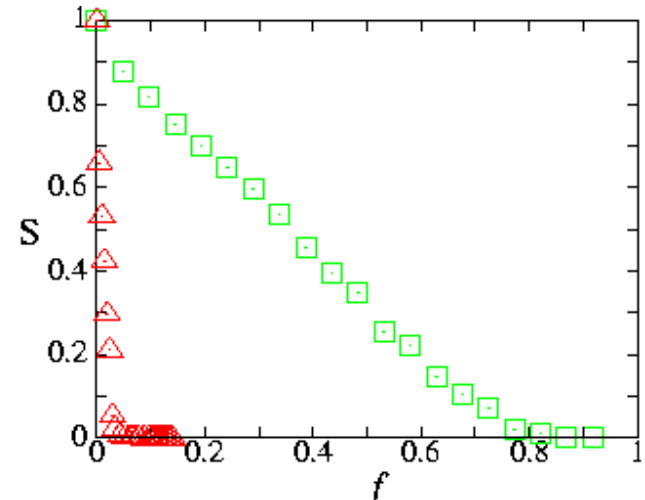
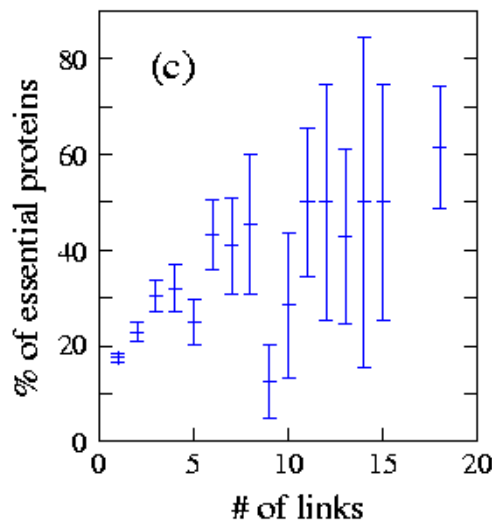
- 5196 gene knockout yeast strains
- Studied growth in rich media and altered environmental conditions
- 19% of genes essential – without them the yeast does not survive even in rich medium
- 15% of knockouts show slow growth in a rich medium
- 15% of strains show morphological alteration – different cell size/shape



# Correlating yeast gene essentiality and protein degree

Start with yeast protein interaction network and knowledge of essential genes. The network topology displays the error tolerant/attack susceptible behavior seen in other networks.

Group proteins by degree, determine the percentage of essential genes (that encode these proteins) in each group.



Green – random node removal  
Red – removal of highest degree node at each step.

Highly connected proteins tend to be more **essential**.

H. Jeong et al., Nature 411, 41 (2001)

# Resilience of metabolic networks

- Nodes – metabolites, edges – reactions
- Gene knockouts – removal of the reaction catalyzed by enzyme
- Consider edge removal (=gene knockout) and node (metabolite) removal
- Determine the lethality fraction of edge or node removal from nodes of given degree
- Relatively narrow range of lethality fraction in case of edge removal
- Very highly connected metabolites are 100% lethal, but...
- The lethality fraction of some less connected nodes is higher than the lethality fraction of more connected nodes.

